PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 7 : C12P 21/00 | A2

(11) International Publication Number: **WO 00/39323**

(43) International Publication Date: 6 July 2000 (06.07.00)

(21) International Application Number: PCT/US99/31010

(22) International Filing Date: 21 December 1999 (21.12.99)

(30) Priority Data:
9828711.3 24 December 1998 (24.12.98) GB

(71) Applicant: GENENCOR INTERNATIONAL, INC. [US/US]; 4 Cambridge Place, 1870 South Winton Road, Rochester, NY 14618 (US).

(71)(72) Applicants and Inventors: DIAZ-TORRES, Maria [ES/US]; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US). FERRARI, Eugenio [-/US]; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304 (US).

(74) Agent: GLAISTER, Debra, J.; Genencor International, Inc., 925 Page Mill Road, Palo Alto, CA 94304-1013 (US).

(81) Designated States: European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

(54) Title: PRODUCTION OF PROTEINS IN GRAM-POSITIVE MICROORGANISMS

(57) Abstract

The present invention provides methods for the production of proteins in gram positive microorganisms, in particular Bacillus species, containing a mutation in at least one of the genes of the opp operon gene cluster.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

Production of Proteins in Gram-Positive Microorganisms

Field of the Invention

The present invention relates to the field of molecular biology and in particular to the production of proteins in gram-positive microorganisms. In particular, the present invention relates to gram-positive microorganisms having a mutation in the opp operon and methods for producing proteins in such host cells.

Background

Gram-positive microorganisms, such as Bacillus, have been used for large-scale industrial fermentation due, in part, to their ability to secrete their fermentation products into the culture media. Secreted proteins are exported across a cell membrane and a cell wall, and then are subsequently released into the external media. It is advantageous to produce proteins of interest in gram-positive microorganisms since exported proteins usually maintain their native conformation.

The opp operon of Bacillus (also known in the art as spoOK operon) encodes an oligopeptide permease that is required for the initiation of sporulation and the development of genetic competence (Rudner et al, 1991, Journal of Bacteriology, 173:1388-1398). The opp operon is a member of the family of ATP-binding cassette transporters involved in the import or export of oligopeptides from 3-5 amino acids. There are five gene products of the opp operon: oppA is the ligand-binding protein and is attached to the outside of the cell by a lipid anchor; oppB and oppC are the membrane proteins that form a complex through which the ligand is transported; oppD and oppF (Perego et al., 1991, Mol. Microbiol. 5:173-185) are the ATPases thought to provide energy for transport (LeDeaux et al., 1997, FEMS Microbiology Letters 153: 63-69). The opp operon has also been referred to as SpoOK by Rudner et al., 1991, J. Bacteriol. 173:1388-1398).

Although deletion mutations in the B.subtilis opp operon have been made (LaDeaux, 1997, FEMS Microbiology Letters 153:63-69) these deletions have not been correlated with enhanced expression of recombinant proteins in B. subtilis. There remains a need for improved methods for the production of proteins in Bacillus as well as other gram-positive microorganisms.

Summary of the Invention

The present invention is based, in part, upon the discovery that a Bacillus strain containing a mutation in the opp operon produces more recombinant protein than the wild-type Bacillus strain. Accordingly, the present invention provides a method for producing a protein in a gram-positive microorganism comprising the steps of obtaining a gram positive microorganism comprising nucleic acid encoding said protein, said microorganism having a mutation in at least one of the genes in the opp operon said mutation resulting in the inactivation of the product of said gene of said opp operon; and culturing said microorganism under conditions suitable for the expression of said protein. In one embodiment of the present invention, the mutation occurs in the oppA gene such that said mutation results in the inactivation of the opp A product. In another embodiment, the gram-positive microorganism is a member of the family *Bacillus*. In another embodiment, the Bacillus includes *B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. circulans, B. lautus and Bacillus thuringiensis*.

In a further embodiment of the present invention, the protein includes hormone, enzyme, growth factor and cytokine. In yet another embodiment, the protein is an enzyme and includes proteases, carbohydrases, and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases and phophatases. In one aspect of the present invention, the protein is protease obtainable from a Bacillus species. In another aspect of the present invention, the protease is subtilisn.

Brief Description of the Drawings

Figures 1A-1M shows the nucleic acid and amino acid sequence of the B.subtilis opp operon. The oppA, oppB, oppC, oppD and oppF genes are designated.

Detailed Description

Definitions

As used herein, the genus *Bacillus* includes all members known to those of skill in the art, including but not limited to *B. subtilis, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. ciculans, B. lautus and B. thuringiensis*.

As used herein the term "Bacillus subtilis opp operon" refers to the B. subtilis operon sequence disclosed in Figures 1A-1M, with the individual genes, oppA, oppB, oppC, oppD and oppF, designated. The term "opp operon" encompasses opp

operons present in gram positive organisms. A gram positive microorganism may have a cluster of multiple genes comprising the opp operon, similar to B. subtilis. The term opp operon refers to the cluster of genes collectively. Gram positive microorganism opp operons are disclosed in Podbielski et al. (1996, Molecular Microbiology 21: 1087-1099 and Tynkkynen et al. (1993, Journal of Bacteriology 175: 7523-7532). Bacillus opp operons will comprise nucleic acid having at least 50%, at least 55%, at least 60%, at least 65%, at least 70%, at least 75%, at least 80%, at least 85%, at least 90% or at least 95% identity to *B.subtilis* opp operon shown in Figure 1 and will function to import peptides for the Bacillus. Percent identity may be determined over the entire length of the opp operon or may be determined on a gene basis for any individual gene in the opp operon gene cluster.

In one embodiment, the gram-positive organisms is a *Bacillus*. In another embodiment, the gram-positive organism is *B. subtilis, B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. ciculans, B. lautus and B. thuringiensis*. In a preferred embodiment, the gram-positive microorganism is *Bacillus subtilis*.

As used herein, "nucleic acid" refers to a nucleotide or polynucleotide sequence, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be double-stranded or single-stranded, whether representing the sense or antisense strand. As used herein "amino acid" refers to peptide or protein sequences or portions thereof.

The terms "isolated" or "purified" as used herein refer to a nucleic acid or amino acid that is removed from at least one component with which it is naturally associated.

As used herein, the term "heterologous protein" refers to a protein or polypeptide that does not naturally occur in a gram-positive host cell. Examples of heterologous proteins include enzymes such as hydrolases including proteases, cellulases, amylases, carbohydrases, and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases and phophatases. The heterologous gene may encode therapeutically significant proteins or peptides, such as growth factors, cytokines, ligands, receptors and inhibitors, as well as vaccines and antibodies. The gene may encode commercially important industrial proteins or peptides, such as proteases, carbohydrases such as amylases and glucoamylases, cellulases, oxidases and lipases. The gene of interest may be a naturally occurring gen , a mutated gene or a synthetic gene.

The term "homologous protein" refers to a protein or polypeptide native or naturally occurring in a gram-positive host cell. The invention includes host cells producing the

homologous protein via recombinant DNA technology. The present invention encompasses a gram-positive host cell having a deletion or interruption of the nucleic acid encoding the naturally occurring homologous protein, such as a protease, and having nucleic acid encoding the homologous protein re-introduced in a recombinant form. In another embodiment, the host cell produces the homologous protein. A recombinant protein refers to any protein encoded by a nucleic acid which has been introduced into the microorganism.

As used herein, the term "mutation" refers to any alteration in at least one of the genes in the opp operon such that the gene product is inactivated or eliminated and transport of oligopeptides of 3-5 amino acids is diminished or eliminated. Examples of mutations include but are not limited to point mutations, frame shift mutations and deletions of part of all of a gene in the opp operon gene cluster. The term "mutation" include alterations in any or all of the genes in the opp operon gene cluster.

Detailed Description of the Preferred Embodiments

The present invention is based upon the discovery that mutating at least one gene of the opp operon in a gram-positive microorganism leads to increased production of heterologous proteins in the mutated microorganism. This discovery provides a basis for producing host microorganisms and expression methods which can be used to produce heterologous proteins. In a preferred embodiment, the host cell is a Bacillus species that has a mutation in oppA such that the oppA gene product is inactivated. The Bacillus is further genetically engineered to produced a heterologous or homologous protein or polypeptide. In one embodiment, the heterologous protein includes proteases, carbohydrases such as amylases and glucoamylases, cellulases, oxidases and lipases. In a preferred embodiment, the polypeptide is a protease obtainable from Bacillus species. In another preferred embodiment, the protease is subtilisn. The nucleic acid and amino acid sequences for subtilisn are found in the following publications: *B. subtilis*: Stahl, M. L. and E. Ferrari 1984 J Bacteriol 158, 411-418; *B. amyloliquefaciens*: Wells, J. A., E. Ferrari, D. J. Henner, D. A. Estell, and E. Y. Chen 1983 Nucleic Acid Res 11, 7911-7925; N. Vasantha, L. D. Thompson, C. Rhodes, et al 1984, J. Bacteriol, 159, 811-819; *B. amylosachariticus*: Kurhara M, Markland, F. S. and Smith E. L. 1972, J. Biol. Chem, 247, 5619-5631; *B. lentus*: Hastrup, S., Branner, S., Norris, F., et al 1989 International Patent No. WO 89/0629; *B. licheniformis*: Jacobs, M., Eliasson, M., Uhlen, M., and Flock, J. 1985, Nucleic Acid Res 13, 8913-8927.

I. Opp operon

The opp operon is known to b associated with the transport, i.e., import of oligopeptides of 3-5 amino acids. The sequence for the Bacillus subtilis opp operon is given in Figures 1A-1M. The present invention encompasses a mutation in at least one of the genes of the opp operon gene cluster such that the gene product is inactivated or eliminated and peptide transport is interrupted. One assay for the presence or absence of a functioning opp operon is to subject the host microorganism to growth in the presence of toxic oligopeptide of 3 amino acids, such as Bialaphos, a tripeptide consiting of two L-alanine molecules and an L-glutamic acid analogue (Meiji Seika, Japan). A microorganism having a functional opp operon will have inhibited growth. A microorganism having a mutation in at least one gene of the opp operon gene cluster will not show growth inhibition in the presence of the toxic oligopeptide.

Gram-positive polynucleotide homologs of *B.subtilis* opp operon may be identified and obtained by standard procedures known in the art from, for example, cloned DNA (*e.g.*, a DNA "library"), genomic DNA libraries, by chemical synthesis once identified, by cDNA cloning, or by the cloning of genomic DNA, or fragments thereof, purified from a desired cell. (*See*, for example, Sambrook *et al.*, 1989, Molecular Cloning, A Laboratory Manual, 2d Ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York; Glover, D.M. (ed.), 1985, DNA Cloning: A Practical Approach, MRL Press, Ltd., Oxford, U.K. Vol. I, II.) The isolated opp operon, or alternatively, individual opp operon genes A, B, C, D, or F, can be molecularly cloned into a suitable vector for propagation. In the molecular cloning of the gene from genomic DNA, DNA fragments are generated, some of which will encode the desired gene. The DNA may be cleaved at specific sites using various restriction enzymes. Alternatively, one may use DNAse in the presence of manganese to fragment the DNA, or the DNA can be physically sheared, as for example, by sonication. The linear DNA fragments can then be separated according to size by standard techniques, including but not limited to, agarose and polyacrylamide gel electrophoresis and column chromatography.

Once the DNA fragments are generated, identification of the specific DNA fragment containing the opp operon may be accomplished in a number of ways. For example, *B.subtilis* oppA gene or its specific RNA, or a fragment thereof, such as a probe or primer, may be isolated and labeled and then used in hybridization assays to detect a gram-positive oppA gene. (Benton, W. and Davis, R., 1977, Science 196:180; Grunstein, M. And Hogness, D., 1975, Proc. Natl. Acad. Sci. USA 72:3961). Those DNA fragments

sharing substantial sequence similarity to the probe will hybridize under stringent conditions.

Hybridization conditions are based on the melting temperature (Tm) of the nucleic acid binding complex, as taught in Berger and Kimmel (1987, Guide to Molecular Cloning Techniques, Methods in Enzymology, Vol 152, Academic Press, San Diego CA) incorporated herein by reference, and confer a defined "stringency" as explained below.

"Maximum stringency" typically occurs at about Tm-5°C (5°C below the Tm of the probe); "high stringency" at about 5°C to 10°C below Tm; "intermediate stringency" at about 10°C to 20°C below Tm; and "low stringency" at about 20°C to 25°C below Tm. As will be understood by those of skill in the art, a maximum stringency hybridization can be used to identify or detect identical polynucleotide sequences while an intermediate or low stringency hybridization can be used to identify or detect polynucleotide sequence homologs.

The term "hybridization" as used herein shall include "the process by which a strand of nucleic acid joins with a complementary strand through base pairing" (Coombs J (1994) Dictionary of Biotechnology, Stockton Press, New York NY).

The process of amplification as carried out in polymerase chain reaction (PCR) technologies is described in Dieffenbach CW and GS Dveksler (1995, PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview NY). A nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides from *B. subtilis* opp operon genes, preferably about 12 to 30 nucleotides, and more preferably about 20-25 nucleotides can be used as a probe or PCR primer.

II. Expression Systems

The present invention provides host microorganisms and expression methods for the production and secretion of desired heterologous proteins in gram-positive microorganisms. In one embodiment, the host cell is genetically engineered to have a mutation in at least one gene of the opp operon gene cluster such that the gene product is eliminated or inactivated. In a preferred embodiment, the mutation is a frame shift mutation in the oppA gene. In another embodiment of the present invention, a gram-positive microorganism having a mutation in at least one gene of the opp operon is genetically engineered to further comprise nucleic acid encoding a heterologous or homologous protein.

Inactivation of genes in the opp operon in a host cell

Producing a gram-positive microorganism incapable of producing at least one gene of the opp operon necessitates inactivating or eliminating the naturally occurring opp operon gene from the genome of the host cell. In a preferred embodiment, the inactivation is the result of a mutation and is preferrably a non-reverting mutation.

One method for mutating nucleic acid encoding a gram-positive microorganism opp operon gene is to clone the nucleic acid or part thereof, modify the nucleic acid by site directed mutagenesis and reintroduce the mutated nucleic acid into the cell on a plasmid. By homologous recombination, the mutated gene may be introduced into the chromosome. In the parent host cell, the result is that the naturally occurring nucleic acid and the mutated nucleic acid are located in tandem on the chromosome. After a second recombination, the modified sequence is left in the chromosome having thereby effectively introduced the mutation into the chromosomal gene for progeny of the parent host cell.

Another method for mutating an opp operon gene is through deleting the chromosomal gene copy. In a preferred embodiment, the entire gene is deleted, the deletion occurring in such as way as to make reversion impossible. In another preferred embodiment, a partial deletion is produced, provided that the nucleic acid sequence left in the chromosome is too short for homologous recombination with a plasmid encoded gene.

Deletion of the naturally occurring gram-positive microorganism opp operon gene can be carried out as follows. An opp gene including its 5' and 3' regions is isolated and inserted into a cloning vector. The coding region of the gene is deleted from the vector *in vitro*, leaving behind a sufficient amount of the 5' and 3' flanking sequences to provide for homologous recombination with the naturally occurring gene in the parent host cell. The vector is then transformed into the gram-positive host cell. The vector integrates into the chromosome via homologous recombination in the flanking regions. This method leads to a gram-positive strain in which the opp gene has been deleted.

The vector used in an integration method is preferably a plasmid. A selectable marker may be included to allow for ease of identification of desired recombinant microorgansims. Additionally, as will be appreciated by one of skill in the art, the vector is preferably one which can be selectively integrated into the chromosome. This can be achieved by introducing an inducible origin of replication, for example, a temperature sensitive origin into the plasmid. By growing the transformants at a temperature to which the origin of replication is sensitive, the replication function of the plasmid is inactivated, thereby providing a means for selection of chromosomal integrants. Integrants may be selected for growth at high temperatures in the presence of the selectable marker, such as an antibiotic. Integration mechanisms are described in WO 88/06623.

Integration by the Campbell-type mechanism can take place in the 5' flanking region of the desired gene, resulting in a strain carrying the entire plasmid vector in the chromosome in the opp locus. Since illegitimate recombination will give different results it will be necessary to determine whether the complete gene has been deleted, such as through nucleic acid sequencing or restriction maps.

Another method of mutating a naturally occurring opp gene is to mutagenize the chromosomal gene copy by transforming a gram-positive microorganism with oligonucleotides which are mutagenic. Alternatively, the chromosomal opp gene can be mutated by replacing the chromosomal gene by a mutant gene by homologous recombination.

The present invention encompasses gram-positive microorganism host cells having additional protease mutations, such as mutations in apr, npr, epr, mpr and others known to those of skill in the art.

Vector Sequences

For production of proteins in a gram-positive microorganism, an expression vector comprising at least one copy of nucleic acid encoding the heterologous or homologous protein, and preferably comprising multiple copies, is transformed into the cell under conditions suitable for expression of the protein. In a preferred embodiment, the protein is a protease obtainable from a Bacillus species.

Expression vectors used in expressing the heterologous proteins of the present invention in gram-positive microorganisms comprise at least one promoter associated with the protein, which promoter is functional in the host cell. In one embodiment of the present invention, the promoter is the wild-type promoter for the selected protein and in another embodiment of the present invention, the promoter is heterologous to the protein, but still functional in the host cell. In one preferred embodiment of the present invention, nucleic acid encoding the protease is stably integrated into the microorganism genome.

In a preferred embodiment, the expression vector contains a multiple cloning site cassette which preferably comprises at least one restriction endonuclease site unique to the vector, to facilitate ease of nucleic acid manipulation. In a preferred embodiment, the vector also comprises one or more selectable markers. As used herein, the term selectable marker refers to a gene capable of expression in the gram-positive host which allows for ease of selection of those hosts containing the vector. Examples of such s lectable markers include but are not limited to antibiotics, such as, erythromycin, actinomycin, chloramphenicol and tetracycline.

III. Transformation

In a preferred embodiment, the host cell is a gram-positive microorganism and in another preferred embodiment, the host cell is *Bacillus*. In one embodiment of the present invention, nucleic acid encoding at least one heterologous protein is introduced into a host cell via an expression vector capable of replicating within the host cell. Suitable replicating plasmids for *Bacillus* are described in Molecular Biological Methods for *Bacillus*, Ed. Harwood and Cutting, John Wiley & Sons, 1990, hereby expressly incorporated by reference; see chapter 3 on plasmids. Suitable replicating plasmids for *B. subtilis* are listed on page 92. Several strategies have been described in the literature for the direct cloning of DNA in *Bacillus*. Plasmid marker rescue transformation involves the uptake of a donor plasmid by competent cells carrying a partially homologous resident plasmid (Contente *et al.*, *Plasmid 2*:555-571 (1979); Haima *et al.*, *Mol. Gen. Genet. 223*:185-191 (1990); Weinrauch *et al.*, *J. Bacteriol. 154(3)*:1077-1087 (1983); and Weinrauch *et al.*, *J. Bacteriol. 169(3)*:1205-1211 (1987)). The incoming donor plasmid recombines with the homologous region of the resident "helper" plasmid in a process that mimics chromosomal transformation.

Transformation by protoplast transformation is described for *B. subtilis* in Chang and Cohen, (1979) Mol. Gen. Genet 168:111-115; for B.megaterium in Vorobjeva et al., (1980) FEMS Microbiol. Letters 7:261-263; for B. amyloliquefaciens in Smith et al., (1986) Appl. and Env. Microbiol. 51:634; for B.thuringiensis in Fisher et al., (1981) Arch. Microbiol. 139:213-217; for B.sphaericus in McDonald (1984) J. Gen. Microbiol. 130:203; and B.larvae in Bakhiet et al., (1985) 49:577. Mann et al., (1986, Current Microbiol. 13:131-135) report on transformation of *Bacillus* protoplasts and Holubova, (1985) Folia Microbiol. 30:97) disclose methods for introducing DNA into protoplasts using DNA containing liposomes. The presence/absence of a marker gene can suggest whether the gene of interest is present in the host microorganism.

Alternatively, host cells which contain the coding sequence for an opp operon gene and express the protein may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridization and protein bioassay or immunoassay techniques which include membrane-based, solution-based, or chip-based technologies for the detection and/or quantification of the nucleic acid or protein.

IV. Assay of Protein Activity

There are various assays known to those of skill in the art for detecting and measuring activity of heterologous proteins or polypeptides. In particular, for proteases, there are assays based upon the release of acid-soluble peptides from casein or hemoglobin measured as absorbance at 280 nm or colorimetrically using the Folin method (Bergmeyer, et al., 1984, Methods of Enzymatic Analysis vol. 5, Peptidases, Proteinases and their Inhibitors, Verlag Chemie, Weinheim). Other assays involve the solubilization of chromogenic substrates (Ward, 1983, Proteinases, in Microbial Enzymes and Biotechnology (W.M. Fogarty, ed.), Applied Science, London, pp. 251-317).

V. Secretion of Recombinant Proteins

Means for determining the levels of secretion of a heterologous or homologous protein in a gram-positive host cell and detecting secreted proteins include, using either polyclonal or monoclonal antibodies specific for the protein. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA) and fluorescent activated cell sorting (FACS). These and other assays are described, among other places, in Hampton R et al (1990, Serological Methods, a Laboratory Manual, APS Press, St Paul MN) and Maddox DE et al (1983, J Exp Med 158:1211).

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting specific polynucleotide sequences include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the nucleotide sequence, or any portion of it, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3 or SP6 and labeled nucleotides.

A number of companies such as Pharmacia Biotech (Piscataway NJ), Promega (Madison WI), and US Biochemical Corp (Cleveland OH) supply commercial kits and protocols for these procedures. Suitable reporter molecules or labels include those radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles and the like. Patents teaching the use of such labels include US Patents 3,817,837; 3,850,752; 3,939,350; 3,996,345; 4,277,437; 4,275,149 and 4,366,241. Also, recombinant immunoglobulins may be produced as shown in US Patent No. 4,816,567 and incorporated herein by reference.

VI. Purification f Pr teins

Gram positive host cells transformed with polynucleotide sequences encoding heterologous or homologous protein may be cultured under conditions suitable for the expression and recovery of the encoded protein from cell culture. The protein produced by a recombinant gram-positive host cell comprising a protease of the present invention will be secreted into the culture media. Other recombinant constructions may join the heterologous or homologous polynucleotide sequences to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins (Kroll DJ et al (1993) DNA Cell Biol 12:441-53).

Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals (Porath J (1992) Protein Expr Purif 3:263-281), protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp, Seattle WA). The inclusion of a cleavable linker sequence such as Factor XA or enterokinase (Invitrogen, San Diego CA) between the purification domain and the heterologous protein can be used to facilitate purification.

VII. Uses of The Present Invention

The present invention provides genetically engineered gram-positive host microorganisms comprising preferably non-revertable mutations in at least one gene of the opp operon gene cluster such that the activity of the gene product is inactivated or eliminated and the transport mechanism is interrupted. The host microorganism may contain additional protease deletions, such as deletions of the mature subtilisn protease and/or mature neutral protease disclosed in United States Patent No. 5,264,366.

In a preferred embodiment, the microorganism is genetically engineered to produce a desired protein or polypeptide. In a preferred embodiment the gram positive microorganism is a *Bacillus* and the polypeptide produced is a protease obtainable from a Bacillus species. In an illustrative embodiment disclosed herein, the protease is subtilisn. In another embodiment disclosed herein, the protein is amylase.

The manner and method of carrying out the present invention may be more fully understood by those of skill in the art by reference to the following examples, which examples are not intended in any manner to limit the scope of the present invention or of the claims directed thereto

Example I

Example I illustrates the increase in production of subtilisn from B. subtilis having a mutation in the oppA gene of the opp operon.

Bacillus subtilis was cultured in the presence of the toxic peptide Bialaphos (50 μg/ml), a tripeptide consiting of two L-alanine molecules and an L-glutamic acid analogue and was shown to have growth inhibition. Bacillus subtilis comprising nucleic acid encoding subtilisn obtainable from B.subtilis and having a mutation in the degU gene (United States Patent No. 5,387,521) did not show inhibition of growth growing on a plate containing Bialaphos (50 μg/ml). This B. subtilis strain was subjected to PCR using the following primers:

Primer 1 GTTGTGGAAACTCAGGTTCATTTGTC and Primer 2 GGCCCGCCCGGTGCTGCTTGC. The PCR fragment generated using the primers was sequenced and found to have a T-insertion at the beginning of the oppA gene.

A plasmid containing a fragment from the oppA wild-type gene was constructed using PCR technique. 857 bp of the *oppA* gene present in Bacillus subtilis was amplified with the addition of restriction sites using the following primers: GCGCGCGGATCCCCTTAAATGATAACTGCTATCAGCGTAAAAACAGGC, introducing BamHI and GCGCGCCTGCAGCACAGCTTTTACTGCCACATCGTCTAGGCTGCC, introducing PstI.

The amplified DNA was cloned in pTSpUCKan plasmid after digesting the plasmid with BamHI and PstI yielding plasmid Pm100. Plasmid pTSpUCKan carries a Kanamycin resistance gene (Kan^r) and a temperature sensitive origin of replication (TsOri). The Kan^r gene was isolated from pJH1, a *Streptococcus faecalis* plasmid (Trieu-Coet, P. and P. Courvalin. 1983. *Gene* 23: 331-341) as a 1.5 kb ClaI fragment. The ClaI fragment was blunt-ended and cloned into the EcoRV site of plasmid pBluescript II KS (Stratagene) to make plasmid pJM114. Plasmid pJM114 was digested with EcoRI and ClaI and the resulting fragment of 1.489 kb was blunt ended, purified and cloned into the NdeI site of plasmid pUC19 (Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Gene 33, 103-119) previously digested with NdeI/blunt -ended and phosphatased yielding plasmid pUS19/Kan. The TsOri was obtained from plasmid pE194 (Villafane, R., Bechhofer, D. H., Narayanan, C. S. and Dubnau, D. (1987), J. Bacteriol. 169: 4822-9; Lovet, P. S. and Ambulus, N. P. Jr (1989) Genetic manipulation of Bacillus subtilis. In Biotechnology Handbooks, Vol 2 (Harwood, C. R., ed.), pp. 115-54. Plenun Press: New York and London) by digestion with the restriction enzyme HinPI. The 1.1 kb fragment was blunt-ended and cloned in the HincII site of plasmid pUC19. This new plasmid, Ts-pUC was

dig sted with XbaI and PstI and the 1.163 Kb fragment was blunt ended, purified and cloned into the blunt-ended AatII site in plasmid pUC19Kan, yielding the final plasmid pTSpUCKan.

The oppA mutant in the B.subtilis strain comprising nucleic acid encoding subtilisn and having a mutation in degU was replaced with the oppA wild-type gene using plasmid pM100. Plasmid pM100 was transformed using standard protoplast transformation techniques (Prapai et al., 1994, Microbiology 140:305) into the B. subtilis strain. Because of the TsOri, this plasmid integrated into the chromosome at the region of homology with the oppA gene when cultured under selective pressure at the non-permissive temperature, e.g. 48°C. After integration, the strain carrying the integrated plasmid was grown extensively at permissive temperature. Upon excision of the integrated plasmid, either the parent strain is restored, or a strain carrying the wild-type gene is constructed. A transformant comprising the wild-type oppA gene was confirmed by nucleic acid sequencing by showing inhibition of growth on Bialaphos.

Example II

Example II illustrates production of subtilisin from strains containing an oppA wild type and an oppA mutant in shake flasks.

A B.subtilis strain comprising nucleic acid encoding B.subtilis subtilisin, a degU mutation and containing wild-type oppA and a strain comprising nucleic acid encoding B.subtilis subtilisin, a degU mutation and containing an oppA mutant produced as in Example I were grown in shake flasks containing 25 ml of LB (Difco) plus 25 ug/mL Chloramphenicol in a 250 mL flask. The shake flasks were incubated at 37^{0}C with vigorous shaking to OD 550 of 0.8. 1 mL of each culture was mixed with 0.5 ml 30 % glycerol and frozen for further experiments. 30 ul of the thawed vials were used to inoculate 40 ml of a media containing 68 g/L Soytone, 300 M PIPES, 20 g/L Glucose (final pH 6.8) in 250 mL flasks. The shake flasks were incubated at 37^{0}C with vigorous shaking for three days, after which aliquots were taken for subtilisin analysis of the supernatant. Results show that the B.subtilis strain containing the oppA mutation produced 19% more subtilisin than when the oppA wild-type gene was present (Table 1).

For the protease assay, supernatants from liquid cultures were harvested after 3 days of growth and assayed for subtilisin as previously described (Estell, D. et al (1985) *J. Biol. Chem.* 260, 6518-6521) in a solution containing 0.3 mM *N*-succinyl-L-Ala-L-Ala-L-Pro-L-Phe-*p*-nitroanalide (Vega Biochemicals), 0.1 M Tris, pH 8.6, at 25^{0}C. The assays

measured the increase in absorbance at 410 nm/min due to hydrolysis and release of *p*-nitroanaline.

Table 1 describes the yields of protease produced from the two strains tested.

Table 1

B. subtilis strain genotype	Subtilisin (g/L)	g/L:OD
oppA-	1.716	18.45
oppAwt	1.39	9.5

Example III

B.subtilis host cells having a T-insertion at the beginning of the oppA gene as described in Example I, were grown in shake flasks containing 25 ml of LB (Difco) in a 250 mL flask. Shake flasks were incubated at 37^{0}C with vigorous shaking and at OD 550 of 0.8, 1 mL of culture was mixed with 0.5 ml 30 % Glycerol and frozen for further experiments. 30 ul of the thawed vials were used to inoculate 40 ml of a media containing 68 g/L Soytone, 300 M PIPES, 20 g/L Glucose (final pH 6.8) in 250 mL flasks. The shake flasks were incubated at 37^{0}C with vigorous shaking for several days during which they were sampled for endogenous amylase activity in the supernatant. Results showed that the strain containing the oppA mutation produced 2 times more B.subtilis endogenous amylase at 48 hours than when the oppA wild-type gene, ie, the wild-type B.subtilis, was present. At 72 hours the increase is 2.4 times, as shown in Table II.

The Amylase Assay

For the amylase assay, whole broth samples were spun down at different times of growth and their supernatants were assayed as follows. 10 ul of the sample was mixed in a cuvete with 790.0 ul of substrate (Megazyme-Ceralpha-Alpha Amylase; substrate is diluted in water and is used as 1 part substrate plus 3 parts of Alpha Amylase buffer pH 6.6) at 25^{0}C. Alpha Amylase buffer is composed of 50mM Maleate Buffer, 5 mM $CaCl^2$, and 0.002 % Triton X-100, PH= 6.7. Amylase was measured in a Spectronic Genesys 2 Spectophotometer using a protocol for amylase activity (Wavelenth: 410 nm, Initial Delay: 75 secs., Total Run Time: 120 secs, Lower Limit: 0.08, Upper Limit: 0.12).

Table 2 describes the yields of amylase produced from the two strains tested.

Table II

B. subtilis strains	Genotype	Amylase (rate) 48 h	60 h
2790	oppAwt	0.054	0.066
2790	oppA-	0.106	0.156

Claims:

1. A method for producing a protein or polypeptide in a gram-positive microorganism comprising the steps of, a) obtaining a gram positive microorganism comprising nucleic acid encoding said protein or polypeptide, said microorganism having a mutation in at least one of the genes in the opp operon said mutation resulting in the inactivation of the product of said gene; and b) culturing said microorganism under conditions suitable for the expression of said protein.

2. The method of Claim 1 wherein said gram-positive microorganism is a member of the family *Bacillus*.

3. The method of Claim 2 wherein said member of the family Bacillus includes *B. licheniformis, B. lentus, B. brevis, B. stearothermophilus, B. alkalophilus, B. amyloliquefaciens, B. coagulans, B. circulans, B. lautus and Bacillus thuringiensis*.

4. The method of Claim 3 wherein said protein includes hormone, enzyme, growth factor and cytokine.

5. The method of Claim 4 wherein said protein is an enzyme.

6. The method of Claim 5 wherein said enzyme includes proteases, carbohydrases, and lipases; isomerases such as racemases, epimerases, tautomerases, or mutases; transferases, kinases and phophatases.

7. The method of Claim 6 wherein said protease is a bacillus protease.

8. The method of Claim 7 wherein said protease is subtilisin.

9. The method of Claim 1 wherein said mutation occurs in the oppA gene such that said mutation results in the inactivation of the opp A product.

10. The method of Claim 1 wherein said mutation is non-revertable.

11. The method of Claim 9 wherein said mutation is a frameshift mutation.

12. The method of Claim 6 wherein said carbohydrase is an amylase.

```
AAGCTTTTCTTTTGATTGCTTCATTATAAATCAATTATAACCAATTGTCATCATGAAAAAACATTCTTTT
                                                                        70
TTCGAAAAGAAAACTAACGAAGTAATATTTAGTTAATATTGGTTAACAGTAGTACTTTTTTGTAAGAAAA

TTCCAGTAAAATTGTAATAATATAAATAACACGAGTGCTGTAAAATCCTTAAATGATAACTGCTATCAGC
                                                                       140
AAGGTCATTTTAACATTATTATATTTATTGTGCTCACGACATTTTAGGAATTTACTATTGACGATAGTCG

GTAAAAACAGGCAGATATTATATGTAAAAAGCAATATGGGCAGAAAACACATGAAAAAGTTTTTAATCAA
                                                                       210
CATTTTTGTCCGTCTATAATATACATTTTTCGTTATACCCGTCTTTTGTGTACTTTTTCAAAAATTAGTT

TTTATGCTTTAAATGGTAGAAGGATATTATGTTCATGGAAGAAAAACTAACGAAGTTTAAATATTTTAAA
                                                                       280
AAATACGAAATTTACCATCTTCCTATAATACAAGTACCTTCTTTTTGATTGCTTCAAATTTATAAAATTT

TTGATAAAATAATATTGCAATAAATTATTTGTTTCATTATAATGAACTTGTTCACTCTATTGTTACAGCT
                                                                       350
AACTATTTTATTATAACGTTATTTAATAAACAAAGTAATATTACTTGAACAAGTGAGATAACAATGTCGA

TTTTTACAAAAATAATCAGAAAAGACGGAACAGAATAAAAGTTGTGGAAACTCAGGTTCATTTGTCTGAT
                                                                       420
AAAAATGTTTTTATTAGTCTTTTCTGCCTTGTCTTATTTTCAACACCTTTGAGTCCAAGTAAACAGACTA

ATTTCTGAGGATTTAGCCGTAAGGAGCTGAAAATTATTATTAGGGGGTTTGCGAATATGAAAAAACGTTG
                                                                       490
TAAAGACTCCTAAATCGGCATTCCTCGACTTTTAATAATAATCCCCCAAACGCTTATACTTTTTTGCAAC

                                        START  Met Lys Lys Arg Trp
                                        oppA

GTCGATTGTCACGTTGATGCTCATTTTCACTCTCGTGCTGAGCGCGTGCGGCTTTGGCGGCAGCGGATCA
                                                                       560
CAGCTAACAGTGCAACTACGAGTAAAAGTGAGAGCACGACTCGCGCACGCCGAAACCGCCGTCGCCTAGT

 Ser Ile Val Thr Leu Met Leu Ile Phe Thr Leu Val Leu Ser Ala Cys Gly Phe Gly Gly Ser Gly Ser
```

FIG._1A

```
AACGGTGAAGGGAAAAAGGACAGTAAAGGAAAGACGACACTTAACATTAATATTAAAACTGAGCCGTTCT
                                                                        630
TTGCCACTTCCCTTTTTCCTGTCATTTCCTTTCTGCTGTGAATTGTAATTATAATTTTGACTCGGCAAGA

 Asn Gly Glu Gly Lys Lys Asp Ser Lys Gly Lys Thr Thr Leu Asn Ile Asn Ile Lys Thr Glu Pro Phe

CCTTACATCCGGGATTGGCAAATGATTCAGTATCAGGCGGTGTTATCCGTCAGACTTTTGAAGGATTGAC
                                                                        700
GGAATGTAGGCCCTAACCGTTTACTAAGTCATAGTCCGCCACAATAGGCAGTCTGAAAACTTCCTAACTG

Ser Leu His Pro Gly Leu Ala Asn Asp Ser Val Ser Gly Gly Val Ile Arg Gln Thr Phe Glu Gly Leu Thr

ACGTATCAATGCAGATGGTGAGCCTGAAGAAGGCATGGCTTCTAAAATTGAAACGAGCAAGGACGGAAAG
                                                                        770
TGCATAGTTACGTCTACCACTCGGACTTCTTCCGTACCGAAGATTTTAACTTTGCTCGTTCCTGCCTTTC

 Arg Ile Asn Ala Asp Gly Glu Pro Glu Glu Gly Met Ala Ser Lys Ile Glu Thr Ser Lys Asp Gly Lys

ACATATACATTTACCATTCGTGATGGTGTGAAATGGTCTAATGGAGACCCTGTAACTGCACAAGATTTTG
                                                                        840
TGTATATGTAAATGGTAAGCACTACCACACTTTACCAGATTACCTCTGGGACATTGACGTGTTCTAAAAC

Thr Tyr Thr Phe Thr Ile Arg Asp Gly Val Lys Trp Ser Asn Gly Asp Pro Val Thr Ala Gln Asp Phe

AATATGCTTGGAAATGGGCGCTTGACCCTAATAATGAATCACAATACGCTTACCAGCTCTACTACATAAA
                                                                        910
TTATACGAACCTTTACCCGCGAACTGGGATTATTACTTAGTGTTATGCGAATGGTCGAGATGATGTATTT

Glu Tyr Ala Trp Lys Trp Ala Leu Asp Pro Asn Asn Glu Ser Gln Tyr Ala Tyr Gln Leu Tyr Tyr Ile Lys

AGGTGCTGAAGCGGCGAATACCGGAAAAGGCAGCCTAGACGATGTGGCAGTAAAAGCTGTGAATGACAAA
                                                                        980
TCCACGACTTCGCCGCTTATGGCCTTTTCCGTCGGATCTGCTACACCGTCATTTTCGACACTTACTGTTT

 Gly Ala Glu Ala Ala Asn Thr Gly Lys Gly Ser Leu Asp Asp Val Ala Val Lys Ala Val Asn Asp Lys

ACGCTGAAGGTTGAATTAAATAACCCGACTCCATATTTCACTGAATTAACTGCGTTCTATACGTATATGC
                                                                       1050
TGCGACTTCCAACTTAATTTATTGGGCTGAGGTATAAAGTGACTTAATTGACGCAAGATATGCATATACG

Thr Leu Lys Val Glu Leu Asn Asn Pro Thr Pro Tyr Phe Thr Glu Leu Thr Ala Phe Tyr Thr Tyr Met
```

FIG._1B

```
CGATCAATAAGAAAATTGCAGAGAAAAATAAAAAGTGGAATACAAATGCCGGAGATGATTATGTATCAAA
                                                                        1120
GCTAGTTATTCTTTTAACGTCTCTTTTTATTTTTCACCTTATGTTTACGGCCTCTACTAATACATAGTTT

Pro Ile Asn Lys Lys Ile Ala Glu Lys Asn Lys Lys Trp Asn Thr Asn Ala Gly Asp Asp Tyr Val Ser Asn

CGGGCCGTTCAAAATGACGGCATGGAAACACAGCGGCTCTATTACTCTCGAAAAAAATGACCAGTATTGG
                                                                        1190
GCCCGGCAAGTTTTACTGCCGTACCTTTGTGTCGCCGAGATAATGAGAGCTTTTTTTTACTGGTCATAACC

Gly Pro Phe Lys Met Thr Ala Trp Lys His Ser Gly Ser Ile Thr Leu Glu Lys Asn Asp Gln Tyr Trp

GATAAAGACAAAGTCAAACTGAAGAAAATCGATATGGTTATGATCAACAATAACAATACGGAACTAAAAA
                                                                        1260
CTATTTCTGTTTCAGTTTGACTTCTTTTAGCTATACCAATACTAGTTGTTATTGTTATGCCTTGATTTTT

Asp Lys Asp Lys Val Lys Leu Lys Lys Ile Asp Met Val Met Ile Asn Asn Asn Asn Thr Glu Leu Lys

AATTCCAAGCTGGCGAACTTGATTGGGCCGGTATGCCGCTCGGACAGCTTCCGACAGAATCCCTGCCGAC
                                                                        1330
TTAAGGTTCGACCGCTTGAACTAACCCGGCCATACGGCGAGCCTGTCGAAGGCTGTCTTAGGGACGGCTG

Lys Phe Gln Ala Gly Glu Leu Asp Trp Ala Gly Met Pro Leu Gly Gln Leu Pro Thr Glu Ser Leu Pro Thr

CCTGAAAAAAGACGGTTCTTTACATGTTGAGCCGATTGCAGGAGTGTATTGGTACAAATTCAACACTGAA
                                                                        1400
GGACTTTTTTCTGCCAAGAAATGTACAACTCGGCTAACGTCCTCACATAACCATGTTTAAGTTGTGACTT

Leu Lys Lys Asp Gly Ser Leu His Val Glu Pro Ile Ala Gly Val Tyr Trp Tyr Lys Phe Asn Thr Glu

GCTAAGCCATTAGACAACGTCAATATCCGTAAAGCTTTAACATATTCGCTTGACCGTCAGTCGATTGTTA
                                                                        1470
CGATTCGGTAATCTGTTGCAGTTATAGGCATTTCGAAATTGTATAAGCGAACTGGCAGTCAGCTAACAAT

Ala Lys Pro Leu Asp Asn Val Asn Ile Arg Lys Ala Leu Thr Tyr Ser Leu Asp Arg Gln Ser Ile Val

AAAACGTTACGCAAGGAGAGCAAATCCCGGCAATGGCTGCAGTGCCGCCTACAATGAAGGGATTTGAGGA
                                                                        1540
TTTTGCAATGCGTTCCTCTCGTTTAGGGCCGTTACCGACGTCACGGCGGATGTTACTTCCCTAAACTCCT

Lys Asn Val Thr Gln Gly Glu Gln Ile Pro Ala Met Ala Ala Val Pro Pro Thr Met Lys Gly Phe Glu Asp
```

FIG._1C

```
TAACAAAGAAGGATACTTCAAAGACAATGATGTCAAAACAGCAAAAGAATACCTTGAAAAAGGCCTAAAA
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1610
ATTGTTTCTTCCTATGAAGTTTCTGTTACTACAGTTTTGTCGTTTTCTTATGGAACTTTTTCCGGATTTT

 Asn Lys Glu Gly Tyr Phe Lys Asp Asn Asp Val Lys Thr Ala Lys Glu Tyr Leu Glu Lys Gly Leu Lys

GAAATGGGCTTAAGCAAGGCATCTGATTTGCCAAAAATCAAATTGTCTTACAACACTGATGACGCACACG
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1680
CTTTACCCGAATTCGTTCCGTAGACTAAACGGTTTTTAGTTTAACAGAATGTTGTGACTACTGCGTGTGC

 Glu Met Gly Leu Ser Lys Ala Ser Asp Leu Pro Lys Ile Lys Leu Ser Tyr Asn Thr Asp Asp Ala His

CGAAAATCGCTCAAGCAGTACAAGAAATGTGGAAGAAAAATTTAGGCGTTGATGTTGAGCTTGATAACTC
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1750
GCTTTTAGCGAGTTCGTCATGTTCTTTACACCTTCTTTTTAAATCCGCAACTACAACTCGAACTATTGAG

Ala Lys Ile Ala Gln Ala Val Gln Glu Met Trp Lys Lys Asn Leu Gly Val Asp Val Glu Leu Asp Asn Ser

AGAGTGGAATGTCTATATTGATAAGCTCCACAGCCAAGATTATCAAATCGGCCGTATGGGCTGGCTCGGC
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1820
TCTCACCTTACAGATATAACTATTCGAGGTGTCGGTTCTAATAGTTTAGCCGGCATACCCGACCGAGCCG

 Glu Trp Asn Val Tyr Ile Asp Lys Leu His Ser Gln Asp Tyr Gln Ile Gly Arg Met Gly Trp Leu Gly

GACTTCAATGATCCTATCAACTTCCTTGAATTGTTCCGCGACAAAAACGGAGGAAATAACGATACAGGCT
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1890
CTGAAGTTACTAGGATAGTTGAAGGAACTTAACAAGGCGCTGTTTTTGCCTCCTTTATTGCTATGTCCGA

 Asp Phe Asn Asp Pro Ile Asn Phe Leu Glu Leu Phe Arg Asp Lys Asn Gly Gly Asn Asn Asp Thr Gly

GGGAAAATCCAGAATTCAAAAAGCTTCTGAATCAGTCACAAACTGAAACAGATAAAACAAAACGTGCAGA
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 1960
CCCTTTTAGGTCTTAAGTTTTTCGAAGACTTAGTCAGTGTTTGACTTTGTCTATTTTGTTTTGCACGTCT

Trp Glu Asn Pro Glu Phe Lys Lys Leu Leu Asn Gln Ser Gln Thr Glu Thr Asp Lys Thr Lys Arg Ala Glu

GCTGCTGAAAAAAGCAGAAGGTATTTTCATTGATGAAATGCCGGTTGCCCCAATCTATTTCTATACTGAT
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+ 2030
CGACGACTTTTTTCGTCTTCCATAAAAGTAACTACTTTACGGCCAACGGGGTTAGATAAAGATATGACTA

 Leu Leu Lys Lys Ala Glu Gly Ile Phe Ile Asp Glu Met Pro Val Ala Pro Ile Tyr Phe Tyr Thr Asp
```

FIG._1D

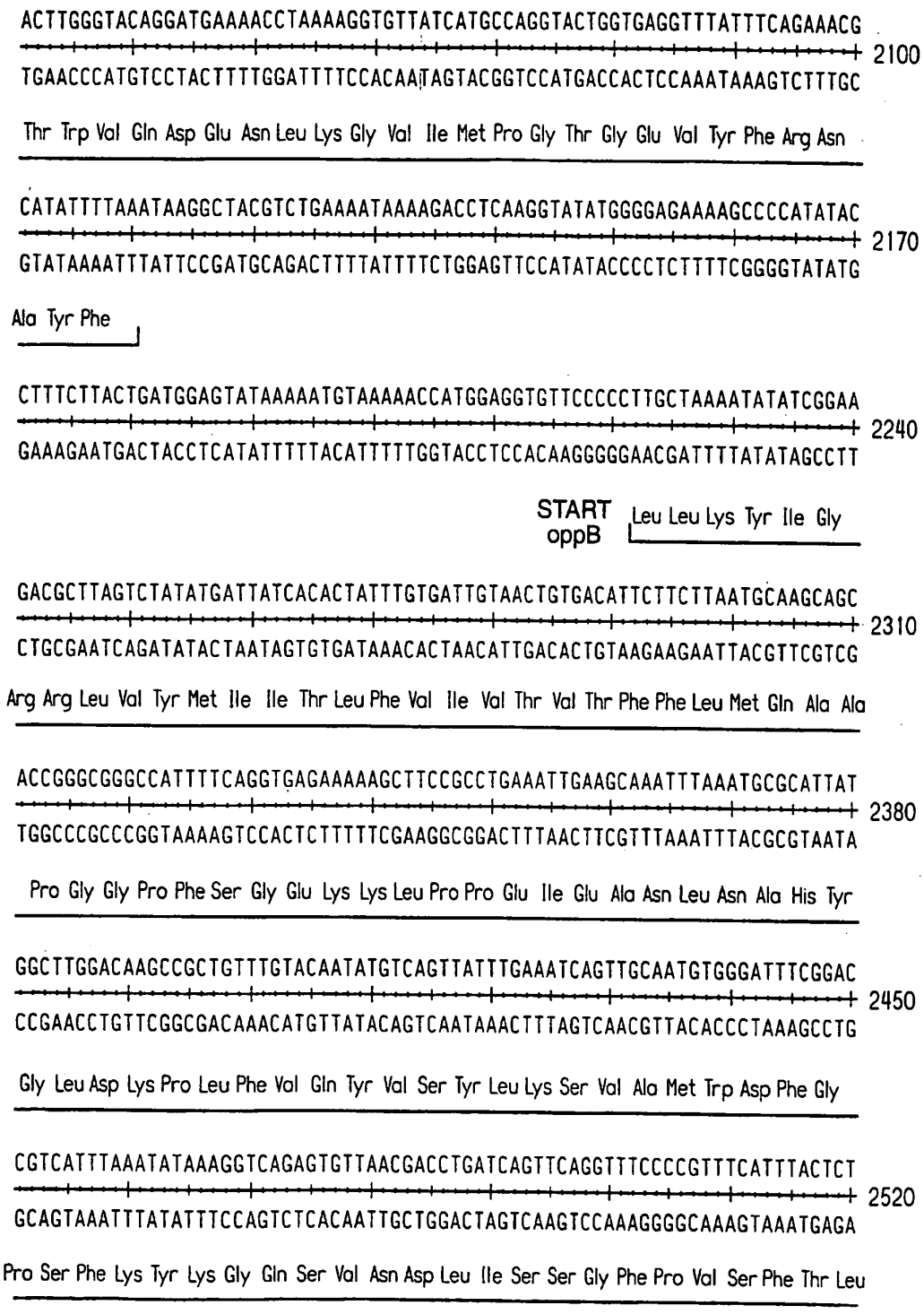
ACTTGGGTACAGGATGAAAACCTAAAAGGTGTTATCATGCCAGGTACTGGTGAGGTTTATTTCAGAAACG
2100
TGAACCCATGTCCTACTTTTGGATTTTCCACAATAGTACGGTCCATGACCACTCCAAATAAAGTCTTTGC
Thr Trp Val Gln Asp Glu Asn Leu Lys Gly Val Ile Met Pro Gly Thr Gly Glu Val Tyr Phe Arg Asn
CATATTTTAAATAAGGCTACGTCTGAAAATAAAAGACCTCAAGGTATATGGGGAGAAAAGCCCCATATAC
2170
GTATAAAATTTATTCCGATGCAGACTTTTATTTTCTGGAGTTCCATATACCCCTCTTTTCGGGGTATATG
Ala Tyr Phe
CTTTCTTACTGATGGAGTATAAAAATGTAAAAACCATGGAGGTGTTCCCCCTTGCTAAAATATATCGGAA
2240
GAAAGAATGACTACCTCATATTTTTACATTTTTGGTACCTCCACAAGGGGGAACGATTTTATATAGCCTT
START
oppB
Leu Leu Lys Tyr Ile Gly
GACGCTTAGTCTATATGATTATCACACTATTTGTGATTGTAACTGTGACATTCTTCTTAATGCAAGCAGC
2310
CTGCGAATCAGATATACTAATAGTGTGATAAACACTAACATTGACACTGTAAGAAGAATTACGTTCGTCG
Arg Arg Leu Val Tyr Met Ile Ile Thr Leu Phe Val Ile Val Thr Val Thr Phe Phe Leu Met Gln Ala Ala
ACCGGGCGGGCCATTTTCAGGTGAGAAAAAGCTTCCGCCTGAAATTGAAGCAAATTTAAATGCGCATTAT
2380
TGGCCCGCCCGGTAAAAGTCCACTCTTTTTCGAAGGCGGACTTTAACTTCGTTTAAATTTACGCGTAATA
Pro Gly Gly Pro Phe Ser Gly Glu Lys Lys Leu Pro Pro Glu Ile Glu Ala Asn Leu Asn Ala His Tyr
GGCTTGGACAAGCCGCTGTTTGTACAATATGTCAGTTATTTGAAATCAGTTGCAATGTGGGATTTCGGAC
2450
CCGAACCTGTTCGGCGACAAACATGTTATACAGTCAATAAACTTTAGTCAACGTTACACCCTAAAGCCTG
Gly Leu Asp Lys Pro Leu Phe Val Gln Tyr Val Ser Tyr Leu Lys Ser Val Ala Met Trp Asp Phe Gly
CGTCATTTAAATATAAAGGTCAGAGTGTTAACGACCTGATCAGTTCAGGTTTCCCCGTTTCATTTACTCT
2520
GCAGTAAATTTATATTTCCAGTCTCACAATTGCTGGACTAGTCAAGTCCAAAGGGGCAAAGTAAATGAGA
Pro Ser Phe Lys Tyr Lys Gly Gln Ser Val Asn Asp Leu Ile Ser Ser Gly Phe Pro Val Ser Phe Thr Leu

FIG._1E

```
TGGAGCAGAAGCTATTCTCCTCGCTTTAGCGTTAGGTGTATTGTTTGGGGTCATTGCAGCCCTTTACCAT
                                                                        2590
ACCTCGTCTTCGATAAGAGGAGCGAAATCGCAATCCACATAACAAACCCCAGTAACGTCGGGAAATGGTA

  Gly  Ala  Glu  Ala  Ile  Leu  Leu  Ala  Leu  Ala  Leu  Gly  Val  Leu  Phe  Gly  Val  Ile  Ala  Ala  Leu  Tyr  His

AATAAGTGGCAGGATTATACCGTCGCGATTTTAACGATATTTGGTATTTCGGTTCCGAGCTTTATCATGG
                                                                        2660
TTATTCACCGTCCTAATATGGCAGCGCTAAAATTGCTATAAACCATAAAGCCAAGGCTCGAAATAGTACC

  Asn  Lys  Trp  Gln  Asp  Tyr  Thr  Val  Ala  Ile  Leu  Thr  Ile  Phe  Gly  Ile  Ser  Val  Pro  Ser  Phe  Ile  Met

CGGCTGTTCTGCAATATGTGTTCTCCATGAAGCTTGGGCTGTTTCCGGTCGCGGGGTGGGATTCCTGGGC
                                                                        2730
GCCGACAAGACGTTATACACAAGAGGTACTTCGAACCCGACAAAGGCCAGCGCCCCACCCTAAGGACCCG

  Ala  Ala  Val  Leu  Gln  Tyr  Val  Phe  Ser  Met  Lys  Leu  Gly  Leu  Phe  Pro  Val  Ala  Gly  Trp  Asp  Ser  Trp  Ala

ATACACCTTTTTGCCTTCCATCGCACTTGCTTCCATGCCGATGGCGTTTATTGCCAGACTTTCCCGTTCA
                                                                        2800
TATGTGGAAAAACGGAAGGTAGCGTGAACGAAGGTACGGCTACCGCAAATAACGGTCTGAAAGGGCAAGT

  Tyr  Thr  Phe  Leu  Pro  Ser  Ile  Ala  Leu  Ala  Ser  Met  Pro  Met  Ala  Phe  Ile  Ala  Arg  Leu  Ser  Arg  Ser

AGCATGATCGAAGTGTTAAACAGTGATTATATCCGCACAGCGAAAGCGAAAGGGCTTTCCGCCCAGCGGT
                                                                        2870
TCGTACTAGCTTCACAATTTGTCACTAATATAGGCGTGTCGCTTTCGCTTTCCCGAAAGGCGGGTCGCCA

  Ser  Met  Ile  Glu  Val  Leu  Asn  Ser  Asp  Tyr  Ile  Arg  Thr  Ala  Lys  Ala  Lys  Gly  Leu  Ser  Ala  Gln  Arg

TACAGTGCGGCACGCCATTCGAAACGCACTTTTGCCGGTTGTTACATATATTGGGCCCGATGGCCGCACA
                                                                        2940
ATGTCACGCCGTGCGGTAAGCTTTGCGTGAAAACGGCCAACAATGTATATAACCCGGGCTACCGGCGTGT

  Leu  Gln  Cys  Gly  Thr  Pro  Phe  Glu  Thr  His  Phe  Cys  Arg  Leu  Leu  His  Ile  Leu  Gly  Pro  Met  Ala  Ala  Gln

GGTCTTAACGGGGAGCTTCATTATTGAAACCATTTTTGGGATTCCGGGGCTTGGTGCACACTTCGTCAAC
                                                                        3010
CCAGAATTGCCCCTCGAAGTAATAACTTTGGTAAAAACCCTAAGGCCCCGAACCACGTGTGAAGCAGTTG

  Val  Leu  Thr  Gly  Ser  Phe  Ile  Ile  Glu  Thr  Ile  Phe  Gly  Ile  Pro  Gly  Leu  Gly  Ala  His  Phe  Val  Asn
```

FIG._1F

```
AGTATTACAAACCGTGATTATACGGTCATTATGGGTGTAACGGTGTTCTTCAGTGTCATCTTGCTATTGT
                                                                        3080
TCATAATGTTTGGCACTAATATGCCAGTAATACCCACATTGCCACAAGAAGTCACAGTAGAACGATAACA

 Ser Ile Thr Asn Arg Asp Tyr Thr Val Ile Met Gly Val Thr Val Phe Phe Ser Val Ile Leu Leu Leu

GTGTATTAATCGTAGATGTGTTATACGGCATTATTGATCCAAGAATCAAGCTTTCCAAAGCAAAGAAAGG
                                                                        3150
CACATAATTAGCATCTACACAATATGCCGTAATAACTAGGTTCTTAGTTCGAAAGGTTTCGTTTCTTTCC

Cys Val Leu Ile Val Asp Val Leu Tyr Gly Ile Ile Asp Pro Arg Ile Lys Leu Ser Lys Ala Lys Lys Gly

AGCCTAGGCCATGCAGAACATTCCAAAAAACATGTTTGAACCAGCCGCAGCGAATGCCGGCGATGCAGAA
                                                                        3220
TCGGATCCGGTACGTCTTGTAAGGTTTTTTGTACAAACTTGGTCGGCGTCGCTTACGGCCGCTACGTCTT

                START oppC
 Ala  .    Met Gln Asn Ile Pro Lys Asn Met Phe Glu Pro Ala Ala Ala Asn Ala Gly Asp Ala Glu

AAAATAAGTAAAAAGAGCCTTTCCCTCTGGAAAGATGCGATGCTTCCGTTCCGCAGCAATAAGCTTGCAA
                                                                        3290
TTTTATTCATTTTTCTCGGAAAGGGAGACCTTTCTACGCTACGAAGGCAAGGCGTCGTTATTCGAACGTT

 Lys Ile Ser Lys Lys Ser Leu Ser Leu Trp Lys Asp Ala Met Leu Pro Phe Arg Ser Asn Lys Leu Ala

TGGTCGGGCTTATCATTATCGTACTTATTATCCTTATGGCAATTTTTGCGCCGATGTTCTCAAGGTATGA
                                                                        3360
ACCAGCCCGAATAGTAATAGCATGAATAATAGGAATACCGTTAAAAACGCGGCTACAAGAGTTCCATACT

Met Val Gly Leu Ile Ile Ile Val Leu Ile Ile Leu Met Ala Ile Phe Ala Pro Met Phe Ser Arg Tyr Asp

TTATTCAACTACTAATCTCTTAAATGCGGATAAGCCGCCTTCAAAAGATCACTGGTTCGGAACAGATGAT
                                                                        3430
AATAAGTTGATGATTAGAGAATTTACGCCTATTCGGCGGAAGTTTTCTAGTGACCAAGCCTTGTCTACTA

 Tyr Ser Thr Thr Asn Leu Leu Asn Ala Asp Lys Pro Pro Ser Lys Asp His Trp Phe Gly Thr Asp Asp

CTTGGACGGGACATTTTCGTCCGTACATGGGTAGGGGCGCGAATCTCAATCTTTATCGGTGTTGCAGCTG
                                                                        3500
GAACCTGCCCTGTAAAAGCAGGCATGTACCCATCCCCGCGCTTAGAGTTAGAAATAGCCACAACGTCGAC

Leu Gly Arg Asp Ile Phe Val Arg Thr Trp Val Gly Ala Arg Ile Ser Ile Phe Ile Gly Val Ala Ala
```

FIG._1G

```
CTGTTCTCGATTTGCTGATCGGCGTCATTTGGGGGAGCATTTCAGGCTTCCGCGGAGGAAGAACAGATGA
                                                                        3570
GACAAGAGCTAAACGACTAGCCGCAGTAAACCCCCTCGTAAAGTCCGAAGGCGCCTCCTTCTTGTCTACT

Ala Val Leu Asp Leu Leu Ile Gly Val Ile Trp Gly Ser Ile Ser Gly Phe Arg Gly Gly Arg Thr Asp Glu

AATCATGATGCGTATCGCTGATATCCTTTGGGCAGTTCCTTCATTATTAATGGTTATCTTACTGATGGTT
                                                                        3640
TTAGTACTACGCATAGCGACTATAGGAAACCCGTCAAGGAAGTAATAATTACCAATAGAATGACTACCAA

 Ile Met Met Arg Ile Ala Asp Ile Leu Trp Ala Val Pro Ser Leu Leu Met Val Ile Leu Leu Met Val

GTTCTTCCGAAAGGTCTATTTACGATTATTATTGCCATGACGATTACAGGCTGGATTAATATGGCCAGAA
                                                                        3710
CAAGAAGGCTTTCCAGATAAATGCTAATAATAACGGTACTGCTAATGTCCGACCTAATTATACCGGTCTT

Val Leu Pro Lys Gly Leu Phe Thr Ile Ile Ile Ala Met Thr Ile Thr Gly Trp Ile Asn Met Ala Arg

TCGTGCGCGGACAAGTGCTGCAGCTGAAGAATCAGGAGTATGTGCTTGCTTCACAGACACTGGGTGCAAA
                                                                        3780
AGCACGCGCCTGTTCACGACGTCGACTTCTTAGTCCTCATACACGAACGAAGTGTCTGTGACCCACGTTT

Ile Val Arg Gly Gln Val Leu Gln Leu Lys Asn Gln Glu Tyr Val Leu Ala Ser Gln Thr Leu Gly Ala Lys

AACATCCCGTCTTCTATTTAAACATATCGTGCCAAACGCAATGGGTTCTATTTTGGTCACGATGACACTG
                                                                        3850
TTGTAGGGCAGAAGATAAATTTGTATAGCACGGTTTGCGTTACCCAAGATAAAACCAGTGCTACTGTGAC

 Thr Ser Arg Leu Leu Phe Lys His Ile Val Pro Asn Ala Met Gly Ser Ile Leu Val Thr Met Thr Leu

ACAGTTCCTACTGCGATTTTTACAGAAGCCTTTTTAAGCTATTTGGGACTTGGTGTTCCGGCTCCGCTGG
                                                                        3920
TGTCAAGGATGACGCTAAAAATGTCTTCGGAAAAATTCGATAAACCCTGAACCACAAGGCCGAGGCGACC

Thr Val Pro Thr Ala Ile Phe Thr Glu Ala Phe Leu Ser Tyr Leu Gly Leu Gly Val Pro Ala Pro Leu

CAAGCTGGGGAACGATGGCTTCTGACGGATTGCCTGCATTGACCTATTATCCGTGGCGTTTATTCTTCCC
                                                                        3990
GTTCGACCCCTTGCTACCGAAGACTGCCTAACGGACGTAACTGGATAATAGGCACCGCAAATAAGAAGGG

Ala Ser Trp Gly Thr Met Ala Ser Asp Gly Leu Pro Ala Leu Thr Tyr Tyr Pro Trp Arg Leu Phe Phe Pro
```

FIG._1H

```
TGCCGGATTTATCTGCATTACAATGTTTGGTTTTAACGTTGTCGGCGACGGATTAAGAGACGCATTGGAT
                                                                        4060
ACGGCCTAAATAGACGTAATGTTACAAACCAAAATTGCAACAGCCGCTGCCTAATTCTCTGCGTAACCTA

  Ala Gly Phe Ile Cys Ile Thr Met Phe Gly Phe Asn Val Val Gly Asp Gly Leu Arg Asp Ala Leu Asp

CCTAAGTTACGTAAATAAGGGAGTGATACGGGTGACACGCCTATTAGAAGTAAAAGATTTAGCAATTTCA
                                                                        4130
GGATTCAATGCATTTATTCCCTCACTATGCCCACTGTGCGGATAATCTTCATTTTCTAAATCGTTAAAGT

                    START oppD
Pro Lys Leu Arg Lys *    Val Ile Arg Val Thr Arg Leu Leu Glu Val Lys Asp Leu Ala Ile Ser

TTTAAAACATATGGCGGAGAGGTCCAGGCGATCCGCGGAGTGAATTTCCATCTGGATAAAGGGGAGACGC
                                                                        4200
AAATTTTGTATACCGCCTCTCCAGGTCCGCTAGGCGCCTCACTTAAAGGTAGACCTATTTCCCCTCTGCG

Phe Lys Thr Tyr Gly Gly Glu Val Gln Ala Ile Arg Gly Val Asn Phe His Leu Asp Lys Gly Glu Thr

TGGCCATTGTTGGAGAATCAGGTTCCGGAAAAAGTGTAACCTCTCAAGCGATTATGAAGCTGATTCCAAT
                                                                        4270
ACCGGTAACAACCTCTTAGTCCAAGGCCTTTTTCACATTGGAGAGTTCGCTAATACTTCGACTAAGGTTA

Leu Ala Ile Val Gly Glu Ser Gly Ser Gly Lys Ser Val Thr Ser Gln Ala Ile Met Lys Leu Ile Pro Met

GCCTCCGGGTTATTTCAAACGCGGTGAGATCCTGTTTGAAGGAAAGGATCTGGTGCCGCTGTCCGAAAAA
                                                                        4340
CGGAGGCCCAATAAAGTTTGCGCCACTCTAGGACAAACTTCCTTTCCTAGACCACGGCGACAGGCTTTTT

  Pro Pro Gly Tyr Phe Lys Arg Gly Glu Ile Leu Phe Glu Gly Lys Asp Leu Val Pro Leu Ser Glu Lys

GAAATGCAAAATGTCCGGGGAAAAGAGATCGGCATGATATTCCAAGATCCGATGACCTCTTTAAATCCAA
                                                                        4410
CTTTACGTTTTACAGGCCCCTTTTCTCTAGCCGTACTATAAGGTTCTAGGCTACTGGAGAAATTTAGGTT

Glu Met Gln Asn Val Arg Gly Lys Glu Ile Gly Met Ile Phe Gln Asp Pro Met Thr Ser Leu Asn Pro

CGATGAAGGTCGGTAAACAAATTACGGAAGTGCTTTTTAAACACGAAAAGATCTCGAAGGAAGCGGCTAA
                                                                        4480
GCTACTTCCAGCCATTTGTTTAATGCCTTCACGAAAAATTTGTGCTTTTCTAGAGCTTCCTTCGCCGATT

Thr Met Lys Val Gly Lys Gln Ile Thr Glu Val Leu Phe Lys His Glu Lys Ile Ser Lys Glu Ala Ala Lys
```

FIG._1I

```
AAAACGCGCGGTTGAACTGCTGGAATTAGTCGGTATCCCAATGCCGGAAAAGCGGGTGAACCAATTTCCG
                                                                        4550
TTTTGCGCGCCAACTTGACGACCTTAATCAGCCATAGGGTTACGGCCTTTTCGCCCACTTGGTTAAAGGC

 Lys Arg Ala Val Glu Leu Leu Glu Leu Val Gly Ile Pro Met Pro Glu Lys Arg Val Asn Gln Phe Pro

CATGAATTTTCAGGCGGGATGAGACAGAGGGTTGTCATTGCCATGGCGCTTGCAGCGAATCCGAAACTTC
                                                                        4620
GTACTTAAAAGTCCGCCCTACTCTGTCTCCCAACAGTAACGGTACCGCGAACGTCGCTTAGGCTTTGAAG

 His Glu Phe Ser Gly Gly Met Arg Gln Arg Val Val Ile Ala Met Ala Leu Ala Ala Asn Pro Lys Leu

TGATCGCCGATGAGCCGACAACTGCTCTTGATGTAACGATTCAAGCGCAAATTTTGGAATTAATGAAGGA
                                                                        4690
ACTAGCGGCTACTCGGCTGTTGACGAGAACTACATTGCTAAGTTCGCGTTTAAAACCTTAATTACTTCCT

Leu Ile Ala Asp Glu Pro Thr Thr Ala Leu Asp Val Thr Ile Gln Ala Gln Ile Leu Glu Leu Met Lys Asp

TTTGCAAAAGAAAATTGACACGTCCATCATCTTTATCACACACGATCTTGGTGTTGTGGCTAACGTTGCT
                                                                        4760
AAACGTTTTCTTTTAACTGTGCAGGTAGTAGAAATAGTGTGTGCTAGAACCACAACACCGATTGCAACGA

 Leu Gln Lys Lys Ile Asp Thr Ser Ile Ile Phe Ile Thr His Asp Leu Gly Val Val Ala Asn Val Ala

GACCGGGTCGCTGTCATGTACGCGGGACAGATTGTAGAAACTGGTACGGTAGACGAAATCTTCTACGACC
                                                                        4830
CTGGCCCAGCGACAGTACATGCGCCCTGTCTAACATCTTTGACCATGCCATCTGCTTTAGAAGATGCTGG

 Asp Arg Val Ala Val Met Tyr Ala Gly Gln Ile Val Glu Thr Gly Thr Val Asp Glu Ile Phe Tyr Asp

CGAGACATCCGTACACTTGGGGGCTTCTTGCATCCATGCCGACACTGGAAAGTTCAGGAGAGGAAGAGCT
                                                                        4900
GCTCTGTAGGCATGTGAACCCCCGAAGAACGTAGGTACGGCTGTGACCTTTCAAGTCCTCTCCTTCTCGA

Pro Arg His Pro Tyr Thr Trp Gly Leu Leu Ala Ser Met Pro Thr Leu Glu Ser Ser Gly Glu Glu Glu Leu

GACTGCAATTCCGGGCACGCCGCCTGATTTGACAAACCCGCCAAAAGGAGATGCTTTTGCCCTGCGGAGC
                                                                        4970
CTGACGTTAAGGCCCGTGCGGCGGACTAAACTGTTTGGGCGGTTTTCCTCTACGAAAACGGGACGCCTCG

 Thr Ala Ile Pro Gly Thr Pro Pro Asp Leu Thr Asn Pro Pro Lys Gly Asp Ala Phe Ala Leu Arg Ser
```

FIG._1J

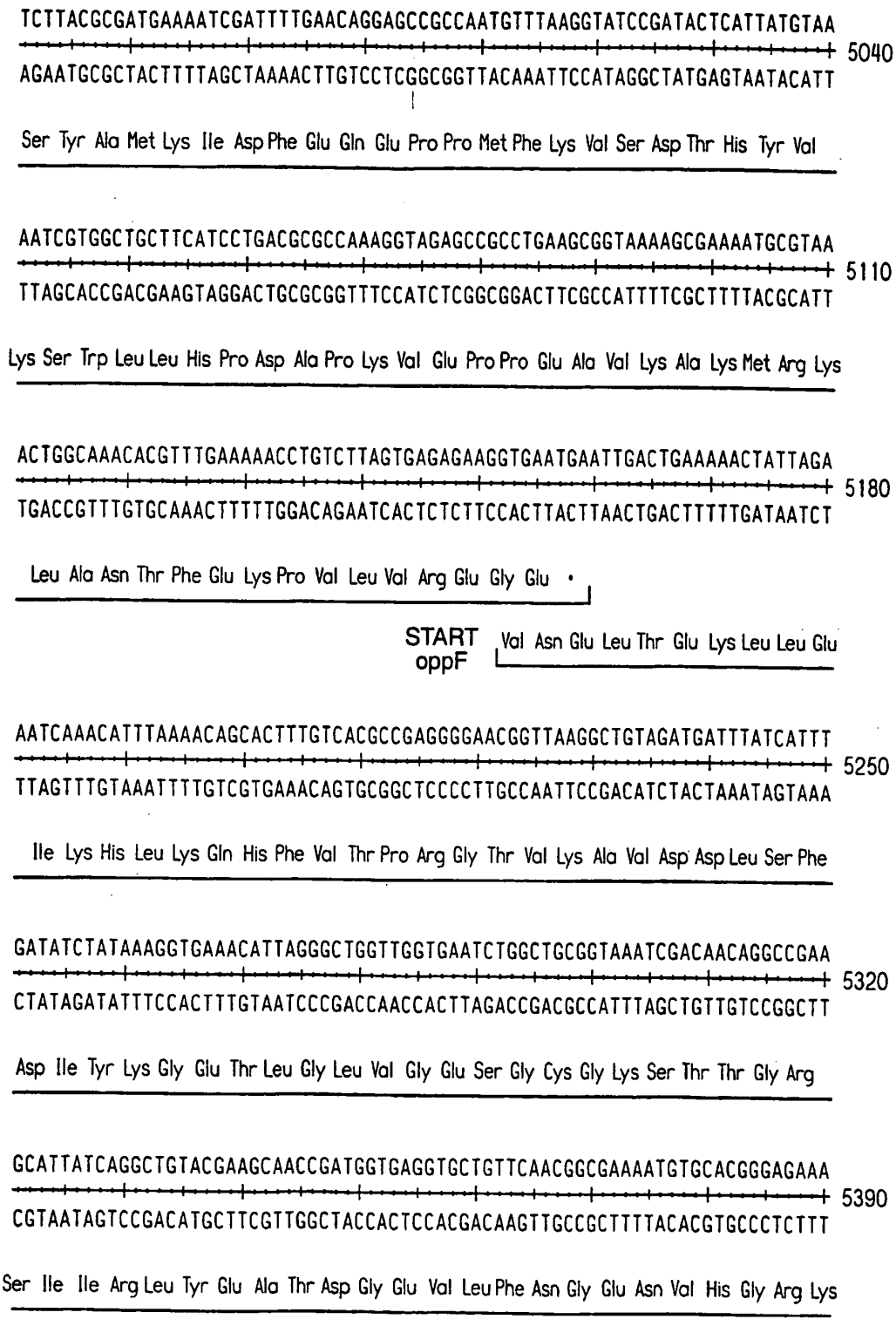
TCTTACGCGATGAAAATCGATTTTGAACAGGAGCCGCCAATGTTTAAGGTATCCGATACTCATTATGTAA
5040
AGAATGCGCTACTTTTAGCTAAAACTTGTCCTCGGCGGTTACAAATTCCATAGGCTATGAGTAATACATT
Ser Tyr Ala Met Lys Ile Asp Phe Glu Gln Glu Pro Pro Met Phe Lys Val Ser Asp Thr His Tyr Val
AATCGTGGCTGCTTCATCCTGACGCGCCAAAGGTAGAGCCGCCTGAAGCGGTAAAAGCGAAAATGCGTAA
5110
TTAGCACCGACGAAGTAGGACTGCGCGGTTTCCATCTCGGCGGACTTCGCCATTTTCGCTTTTACGCATT
Lys Ser Trp Leu Leu His Pro Asp Ala Pro Lys Val Glu Pro Pro Glu Ala Val Lys Ala Lys Met Arg Lys
ACTGGCAAACACGTTTGAAAAACCTGTCTTAGTGAGAGAAGGTGAATGAATTGACTGAAAAACTATTAGA
5180
TGACCGTTTGTGCAAACTTTTTGGACAGAATCACTCTCTTCCACTTACTTAACTGACTTTTTGATAATCT
Leu Ala Asn Thr Phe Glu Lys Pro Val Leu Val Arg Glu Gly Glu
START
oppF
Val Asn Glu Leu Thr Glu Lys Leu Leu Glu
AATCAAACATTTAAAACAGCACTTTGTCACGCCGAGGGGAACGGTTAAGGCTGTAGATGATTTATCATTT
5250
TTAGTTTGTAAATTTTGTCGTGAAACAGTGCGGCTCCCCTTGCCAATTCCGACATCTACTAAATAGTAAA
Ile Lys His Leu Lys Gln His Phe Val Thr Pro Arg Gly Thr Val Lys Ala Val Asp Asp Leu Ser Phe
GATATCTATAAAGGTGAAACATTAGGGCTGGTTGGTGAATCTGGCTGCGGTAAATCGACAACAGGCCGAA
5320
CTATAGATATTTCCACTTTGTAATCCCGACCAACCACTTAGACCGACGCCATTTAGCTGTTGTCCGGCTT
Asp Ile Tyr Lys Gly Glu Thr Leu Gly Leu Val Gly Glu Ser Gly Cys Gly Lys Ser Thr Thr Gly Arg
GCATTATCAGGCTGTACGAAGCAACCGATGGTGAGGTGCTGTTCAACGGCGAAAATGTGCACGGGAGAAA
5390
CGTAATAGTCCGACATGCTTCGTTGGCTACCACTCCACGACAAGTTGCCGCTTTTACACGTGCCCTCTTT
Ser Ile Ile Arg Leu Tyr Glu Ala Thr Asp Gly Glu Val Leu Phe Asn Gly Glu Asn Val His Gly Arg Lys

FIG._1K

ATCGCGGAAAAAGCTGCTGGAATTCAACCGCAAAATGCAGATGATTTTCCAAGACCCCTATGCATCCCTG
TAGCGCCTTTTTCGACGACCTTAAGTTGGCGTTTTACGTCTACTAAAAGGTTCTGGGGATACGTAGGGAC 5460

Ser Arg Lys Lys Leu Leu Glu Phe Asn Arg Lys Met Gln Met Ile Phe Gln Asp Pro Tyr Ala Ser Leu

AATCCGAGAATGACAGTTGCTGATATTATTGCTGAAGGCCTTGATATTCATAAGCTGGCAAAAACGAAAA
TTAGGCTCTTACTGTCAACGACTATAATAACGACTTCCGGAACTATAAGTATTCGACCGTTTTTGCTTTT 5530

Asn Pro Arg Met Thr Val Ala Asp Ile Ile Ala Glu Gly Leu Asp Ile His Lys Leu Ala Lys Thr Lys

AAGAGCGGATGCAGCGAGTTCATGAACTATTGGAAACAGTGGGATTGAACAAGGAACACGCGAACCGCTA
TTCTCGCCTACGTCGCTCAAGTACTTGATAACCTTTGTCACCCTAACTTGTTCCTTGTGCGCTTGGCGAT 5600

Lys Glu Arg Met Gln Arg Val His Glu Leu Leu Glu Thr Val Gly Leu Asn Lys Glu His Ala Asn Arg Tyr

TCCTCATGAATTTTCCGGCGGCCAGCGCCAAAGAATCGGGATTGCCAGAGCGCTTGCTGTTGATCCGGAA
AGGAGTACTTAAAAGGCCGCCGGTCGCGGTTTCTTAGCCCTAACGGTCTCGCGAACGACAACTAGGCCTT 5670

Pro His Glu Phe Ser Gly Gly Gln Arg Gln Arg Ile Gly Ile Ala Arg Ala Leu Ala Val Asp Pro Glu

TTCATTATCGCGGATGAGCCGATTTCCGCTTTGGATGTATCCATTCAAGCGCAGGTCGTGAATTTAATGA
AAGTAATAGCGCCTACTCGGCTAAAGGCGAAACCTACATAGGTAAGTTCGCGTCCAGCACTTAAATTACT 5740

Phe Ile Ile Ala Asp Glu Pro Ile Ser Ala Leu Asp Val Ser Ile Gln Ala Gln Val Val Asn Leu Met

AAGAACTGCAAAAAGAAAAAGGGCTCACATACCTGTTTATTGCCCACGATTTATCGATGGTCAAATACAT
TTCTTGACGTTTTTCTTTTTCCCGAGTGTATGGACAAATAACGGGTGCTAAATAGCTACCAGTTTATGTA 5810

Lys Glu Leu Gln Lys Glu Lys Gly Leu Thr Tyr Leu Phe Ile Ala His Asp Leu Ser Met Val Lys Tyr Ile

CAGTGACCGCATTGGCGTCATGTATTTCGGGAAACTGGTTGAGCTTGCGCCGGCAGATGAGCTTTATGAA
GTCACTGGCGTAACCGCAGTACATAAAGCCCTTTGACCAACTCGAACGCGGCCGTCTACTCGAAATACTT 5880

Ser Asp Arg Ile Gly Val Met Tyr Phe Gly Lys Leu Val Glu Leu Ala Pro Ala Asp Glu Leu Tyr Glu

FIG._1L

```
AATCCGCTTCACCCATATACAAAATCATTGCTTTCTGCGATTCCGCTTCCTGATCCGGACTATGAGAGAA
                                                                        5950
TTAGGCGAAGTGGGTATATGTTTTAGTAACGAAAGACGCTAAGGCGAAGGACTAGGCCTGATACTCTCTT

Asn Pro Leu His Pro Tyr Thr Lys Ser Leu Leu Ser Ala Ile Pro Leu Pro Asp Pro Asp Tyr Glu Arg

ATCGCGTTCGCCAGAAATATGATCCGTCTGTCCATCAATTAAAGGATGGGGAAACGATGGAATTCCGTGA
                                                                        6020
TAGCGCAAGCGGTCTTTATACTAGGCAGACAGGTAGTTAATTTCCTACCCCTTTGCTACCTTAAGGCACT

Asn Arg Val Arg Gln Lys Tyr Asp Pro Ser Val His Gln Leu Lys Asp Gly Glu Thr Met Glu Phe Arg Glu

AGTCAAACCGGGACATTTTGTGATGTGCACGGAAGCCGAATTTAAAGCTTTTTCATGATTCATCAATCCT
                                                                        6090
TCAGTTTGGCCCTGTAAAACACTACACGTGCCTTCGGCTTAAATTTCGAAAAAGTACTAAGTAGTTAGGA

Val Lys Pro Gly His Phe Val Met Cys Thr Glu Ala Glu Phe Lys Ala Phe Ser •

TCAAGAGATTTCTCTTGAAGGATTTTTTTGCGTCTTCATAGAAAGTGAGAATGATAACATTTACAATTAG
                                                                        6160
AGTTCTCTAAAGAGAACTTCCTAAAAAAACGCAGAAGTATCTTTCACTCTTACTATTGTAAATGTTAATC

AGGAAAAAGCGGAGGCGAAATACGATTCAATTTCTGCATAACAAAAATGTTTTTGCGCTATTGCTGTCTC
                                                                        6230
TCCTTTTTCGCCTCCGCTTTATGCTAAGTTAAAGACGTATTGTTTTTACAAAAACGCGATAACGACAGAG

AATCTTTTCAATCTTTAGCAGGAGTCCTCGTCACAATCGTTTTAATGGTCCGGATCTATCAAATGACTGA
                                                                        6300
TTAGAAAAGTTAGAAATCGTCCTCAGGAGCAGTGTTAGCAAAATTACCAGGCCTAGATAGTTTACTGACT
```

FIG._1M